Unsupervised Environmental Sound Classification Based On Topological Persistence

Yueqi Cao[1], Shiqiang Zhang[1], Fangjia Yan[1], Wenyu Li[2], Fupeng Sun[1], Huafei Sun[1*],
[1]School of Mathematics and Statistics, Beijing Institute of Technology, Beijing, China
[2]School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China
[*]Email address: huafeisun@bit.edu.cn

Environmental sound classification (ESC) is a significant problem in signal processing and machine learning. In this paper, a novel unsupervised learning approach is proposed using topological data analysis. Each sound signal is transformed into a point cloud by time-delay embedding. The persistence diagrams are obtained after computing the persistent homology of Vietoris-Rips filtrations. The sound signals are assigned to different classes using kmeans algorithm based on the bottleneck distance matrix of persistence diagrams.

***Index Terms*—Environmental sound classification, Topological data analysis, Time-delay embedding, Unsupervised learning**

I. Introduction

The problem of environmental sound classification (ESC) has been attracting the interests of scientists from various fields. On the one hand, ESC is instrumental in practice and has many applications, such as robot navigation and keyword-based audio retrieval [2], [3]. On the other hand, with the rapid development of machine learning and pattern recognition, several learning methods are being proposed to solve this classification problem [5], [7]. However, though much progress has been made on ESC in past decades, the problem has not been perfectly solved. As pointed out in [1], environment sounds (like animal sounds, thunder, rain etc.) are not similar to any other type of audio signals. Unlike speech recognition, it is not likely to find sub-structures such as phonemes in ESC. Compared with music, ESC has no meaningful stationary patterns such as melody and rhythm. Therefore, the problem of ESC still needs new aspects and algorithms.

Topological data analysis (TDA) is a powerful tool to deal with data which is great in both size and complexity. Recent years have witnessed the successes of TDA applied in natural image statistics [8], [9], shape modelling [10], [12], chaotic signals [6], [11] and etc. Compared with traditional methods in data analysis, one attempts to find coarser and qualitative global features in computational topology. The concept of connectedness is generalized to any dimension, known as homology or Betti numbers. In the framework of TDA, a nested combinatorial object, called filtration, is built on the distance or similarity information of data sets. Then, one applies a standard algorithm to compute the persistent homology with respect to the filtration. The topological features are extracted and displayed in the form of bar codes or multisets, which are called persistence diagrams. TDA is useful because of its robustness to noises. Novel stability theorems are presented in [12], [13].

In this paper, we apply TDA to the problem of ESC. There are related works such as time series classification, wheeze detection and chaotic dynamic systems based on TDA. In [14], the authors examined the end-to-end TDA processing pipeline for persistent homology applied to time-delay embeddings of time series. In [15], TDA is used to find the harmonic structure of a periodic signal. The sound signals are transformed into point clouds using time-delay embedding. Statistical inference are set up after obtaining persistence bar codes. Experiments on breathing sound data show that this method is valid. In [16], chaotic time series are considered. TDA is used to extract the structure of attractors. Similarly, we use time-delay embedding to convert a environment sound into a point cloud. However, we define the eccentricity function to apply a different and efficient subsampling algorithm. Based on the bottleneck distances of persistence diagrams, we apply the multidimensional scaling (MDS) algorithm to visualize the persistence diagrams and use standard kmeans algorithm to classify them. This yields an unsupervised learning method to ESC. Experiments show that our method is efficient.

The rest of this paper is organized as follows. In section II, we sketch the basic concepts in persistent homology. Details of the classification algorithm are presented in section III. In section IV, we apply our algorithm to real data. We conclude this paper with a discussion on future work in section V.

II. Topological Backgrounds

The concept of persistent homology starts with the classic notion of simplicial homology groups. Let K be a simplicial complex. The simplices in various dimensions span the free abelian groups $\{C_p(K)\}$. A series of homomorphisms $\partial_p : C_{p+1} \to C_p(K)$ satisfy the rule $\partial_p\partial_{p+1} = 0$. Therefore, the group $\mathrm{im}\partial_{p+1}$ is a subgroup of $\ker\partial_p$. The pth homology group $H_p(K)$ is defined to be the quotient $\ker\partial_p/\mathrm{im}\partial_{p+1}$.

Assume there is a nested sequence of simplicial complexes

$$\emptyset = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n. \tag{1}$$

The sequence (1) is called a filtration. A filtration gives rise to a sequence of homomorphisms for homology in any dimension p

$$0 = H_p(K_0) \to H_p(K_1) \to \cdots \to H_p(K_n). \tag{2}$$

978-1-7281-2345-5/19/$31.00 ©2019 IEEE

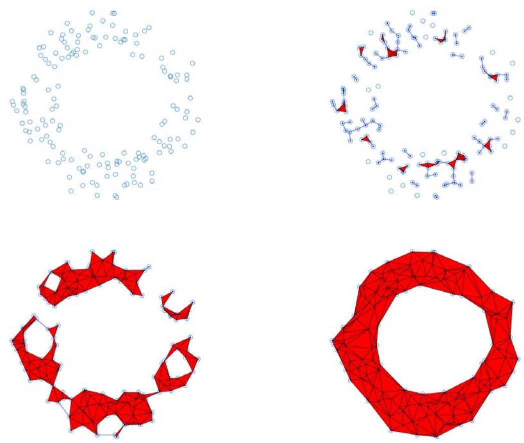

Fig. 1. A filtration based on 100 samples from a noisy circle.

Denote the homomorphism between $H_p(K_i)$ and $H_p(K_j)$ by $f_p^{i,j} : H_p(K_i) \to H_p(K_j), i \leq j$. The image of $f_p^{i,j}$ is called a p-th persistent homology group. Suppose $[\sigma]$ is a nontrivial element in $H_p(K_i)$. If $f_p^{i,j-1}([\sigma]) \neq 0, f_p^{i,j}([\sigma]) = 0$, we say $[\sigma]$ dies entering K_j. Furthermore, if $[\sigma]$ is not the image of any homomorphism $f_p^{l,i}, l < i$, $[\sigma]$ is said to be born in K_i. From this point of view, we see that the persistent homology group $f_p^{i,j}(H_p(K_i))$ contains elements that persist between K_i and K_j.

The pth homology classes are encoded by points (r_i, r_j), where r_i is the birth time and r_j is the death time. The collection of these points, denoted by $\mathcal{D}_p$, is a multiset in the extended plane $\bar{\mathbb{R}}^2$. $\mathcal{D}_p$ is called the pth persistence diagram. Persistence diagrams extract all the information for a filtration by a structure theorem stated by A. Zomorodian and G. Carlsson [17].

Given a topological space X, it is natural to ask a way to construct simplicial complexes based on this space. A popular technique is use covers. Suppose $\mathcal{U} = \{U_i\}_{i=1}^n$ is a cover of X. Set each U_i to be a vectex. A k-simplex is spanned by $k+1$ elements in $\mathcal{U}$ if the $k+1$-intersection of these sets is not empty. The corresponding simplicial complex is called the nerve of $\mathcal{U}$. Under mild assumptions on X and $\mathcal{U}$, we can obtain important topological information from nerve. For example, if X is a subset in Euclidean space $\mathbb{R}^d$ and $\mathcal{U}$ is a collection of closed, convex sets, the following theorem states the homology of nerve and X are the same.

Theorem ([18]). *Let $\mathcal{U}$ be a finite collection of closed, convex sets in Euclidean space. Then the nerve of $\mathcal{U}$ and the union of the sets in $mathcalU$ have the same homotopy type.*

This construction is generalized to filtration. Let S be a finite set of points in $\mathbb{R}^d$ and let $B_x(r)$ be the closed ball of radius r centered at $x \in S$. The Čech complex of S with parameter r, denoted by $\check{C}ech(S, r)$, is defined to the nerve of the cover $\{B_x(r) | x \in S\}$. Note that when $r < r'$, then $\check{C}ech(S, r) \subseteq \check{C}ech(S, r')$. Let the parameter vary. We obtain what is called Čech filtration.

There are many constructions of filtration based on finite metric spaces other than Čech filtration. For example, Vietoris-Rips filtration is simplified version of Čech filtration. We will discuss the details in algorithm description.

III. Classification Algorithm

The framework of our classification algorithm can be summarized as follows. Given a sound signal, we transform it into a point cloud using time-delay embedding. The delay is chosen based on a time-varying autocorrelation function. Then we define an eccentric function and sample a subset of the point cloud according to the eccentricity. Next we construct a Vietoris-Rips filtration based on the subsample and compute its persistence diagrams. Finally, we compute the bottleneck distance between persistence diagrams and apply the standard kmeans algorithm to obtain the clustering.

A. Time-delay Embedding

Time-delay embedding method was firstly proposed in [19], [20]. It embeds a scalar series into a d dimensional Euclidean space. More specifically, let $f = \{f_i\}_{i=1}^n$ be a signal of length n. It is mapped into a point set $S_{\tau,d}(f) = \{s_k\}_{k=1}^{n'}$ in $\mathbb{R}^d$. Here

$$s_k = [f_k, f_{k+\tau}, \cdots, f_{k+(d-1)\tau}] \in \mathbb{R}^d, \tag{3}$$

and the number of $S_{\tau,d}(f)$ is

$$n' = n - (d-1)\tau. \tag{4}$$

where τ is called the sampling lag and d is called the embedding dimension. Note that time-delay embedding partite the signal into segmentations. This technique is often used to transform a time series into the phase space. One needs to be careful with the choice of delay τ. If τ is too small, s_k will be dependent on each other. If τ is too large, different s_k will be irrelevant. This problem is solved by considering the following autocorrelation function

$$R(t) = \sum_i f_{i+t} f_i, \tag{5}$$

Then select $t_{c1} < \tau < t_{c2}$, where t_{c1} and t_{c2} are the first and second critical values of $R(t)$ respectively. As shown in [21], [22], this choice of τ guarantees a reasonable delay and the geometry of the point cloud is closely related to the original signal.

B. Subsampling

In general, the point cloud obtained from time-delay embedding is in large size. It is crucial to sample a subset to manipulate the point cloud. Meanwhile, the topological feature, as well as the signal feature, should be kept as much as possible. A simple observation is that, after time-delay embedding, most points lie in a small region. Since for a environment sound signal, most time we are hearing the noise from the environment, while the true feature is usually short and quick. This observation inspired us to sample the points

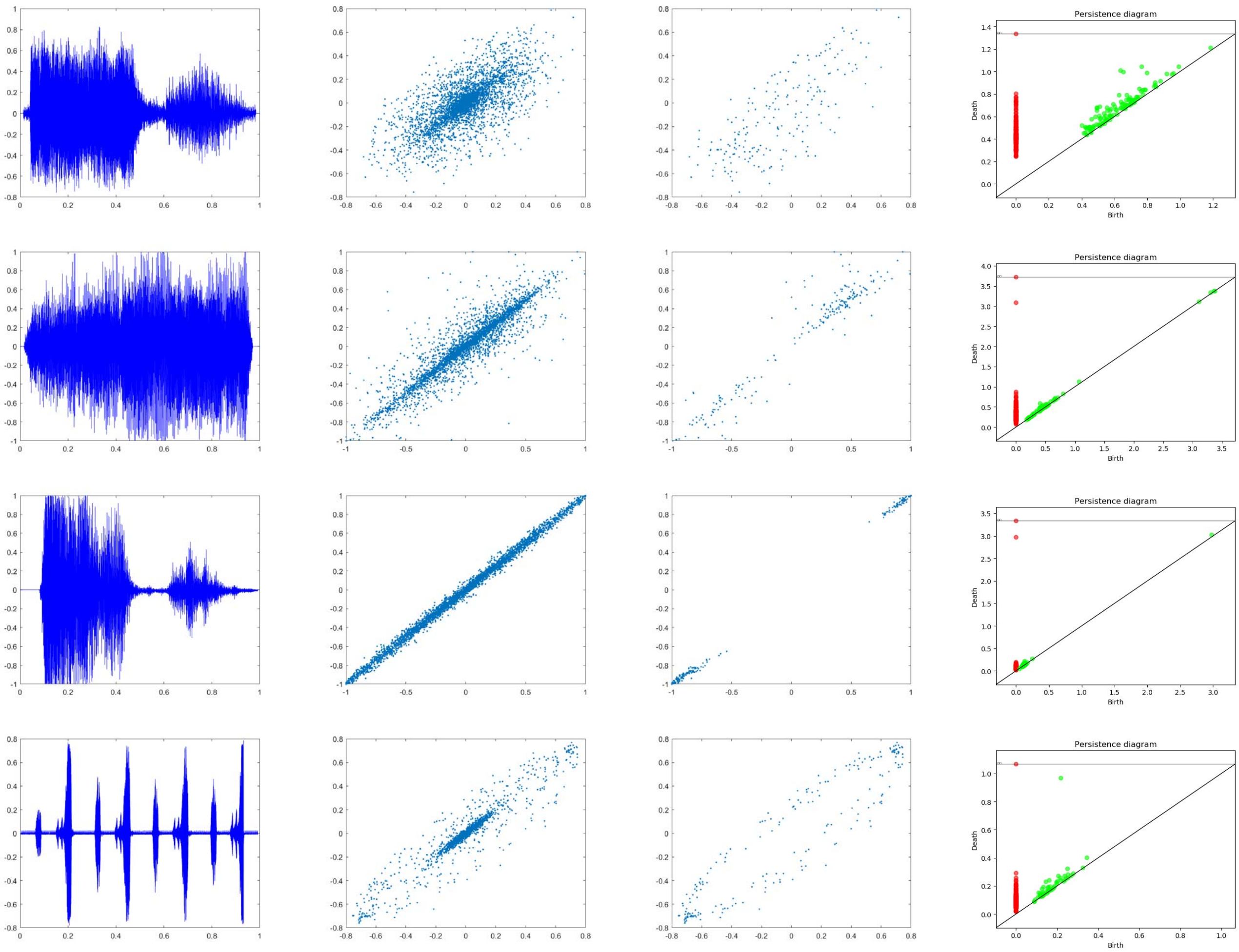

Fig. 2. The figures from the top row to the bottom row present the cases of elephant, bee, lion and nightingale respectively. The first column presents the original signals of animal sounds. The second column presents the projections of embedding point clouds to the first two coordinates. Subsamples using eccentricity are presented in the third column. The last column contains the persistence diagrams of point clouds.

with low density. Furthermore, this sample scheme is required to keep the stable property of persistence. To achieve this, we need to sample using a 'proper' function, as discussed in [12]. Let $\mathbb{H}$ be the set of all finite metric spaces, and let $(X, d_X), (Y, d_Y) \in \mathbb{H}$ be two distinct finite metric spaces. A correspondence between X and Y is a subset $C \subseteq X \times Y$ such that for any $x \in X$, there is $y \in Y$ with $(x, y) \in C$ and for any $y \in Y$, there is $x \in X$ with $(x, y) \in C$. A class of functions $\{f_X(\cdot)\}_{X \in \mathbb{H}}$ is called admissible if for any $(X, d_X) \in \mathbb{H}$, $f_X(\cdot) : X \to \mathbb{R}$ is a continuous function and for distinct X and Y

$$||f_X - f_Y||_{l^\infty(C)} \leq \frac{L}{2} \sup_{(x,y),(x',y') \in C} |d_X(x, x') - d_Y(y, y')|, \tag{6}$$

where L is a constant and C is any correspondence. Admissible functions guarantee that the filtration with respect to level sets of functions have stable persistence under Gromov-Hausdorff distance. As shown in [12], the q order eccentricity map is admissible. For any finite metric space X, the q order eccentricity is defined as

$$e_X^q(x_0) = \max_{x_1, \cdots, x_q} \min_{0 \leq i < j \leq q} d_X(x_i, x_j), \tag{7}$$

for any $x_0 \in X$. For the special case where $q = 1$, $e_X^1(x_0) = \max_{x \in X} d_X(x_0, x)$. We write e_X for short. Note that $e_X(x_0)$ finds the maximal distance of x_0 in X. Therefore, it measures how a point stays far away from other points in the whole set.

Eccentricities are suitable for us to propose a sampling scheme. On the one hand, the level sets have stable persistence. On the other hand, points with large eccentricities correspond the features in a sound signal. Suppose S is a point cloud obtained from time-delay embedding, we compute the eccentricities for all points in S. Then we preserve the points with eccentricities lager than δ, where δ is a preset threshold.

C. Vietoris-Rips Filtration

In section II, we have shown how to construct a filtration based on a finite metric space. That is, we cover the space using closed balls and take the nerve of this cover. However,

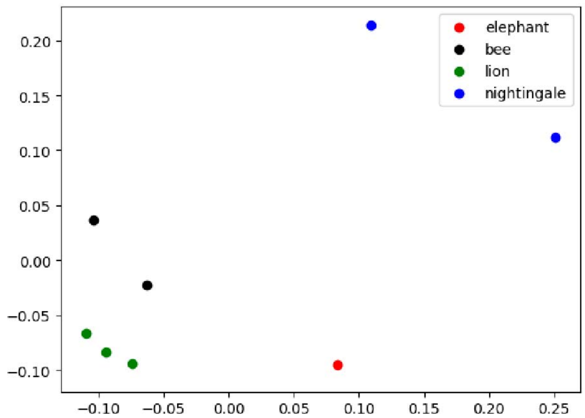

Fig. 3. Clustering results of animal sounds.

it is inconvenient to check the intersections of balls, especially when we want to construct a simplicial complex of high dimension. Therefore, we use a simplified version of Čech complex called Vietoris-Rips complex.

Let $(X, d_X) \in \mathbb{H}$ be a finite metric space and $r > 0$ be a preset parameter. The Vietoris-Rips complex $VR(X, r)$ is defined as the simplicial complex such that

- The vertices are points in X;
- $\{x_1, x_2, \cdots, x_m\}$ spans an $m-1$ dimensional simplex if $\max_{i,j} d_X(x_i, x_j) \leq 2r$.

It is clear from definition that $VR(X, r) \subseteq VR(X, r')$ if $r < r'$. Let the parameter r vary and one obtains a filtration. Note that the second condition implies a simplex is in $VR(X, r)$ if and only if all its edges are in $VR(X, r)$. Hence, the Vietoris-Rips complex is completely determined by its one dimensional skeleton. Suppose X is a subset of Euclidean space. An edge $\{x_i, x_j\}$ is in $VR(X, r)$ if and only if $||x_i - x_j|| \leq 2r$, which is equivalent to $B_{x_i}(r) \cap B_{x_j}(r) \neq \emptyset$. From this point of view we see that $\check{C}ech(X, r) \subseteq VR(X, r)$ in Euclidean space. Furthermore, if we use ∞-norm for $\mathbb{R}^d$, then the two complexes are equal [23]. Under this condition the nerve theorem implies no topological information is lost using Vietoris-Rips complex.

D. Bottleneck Distance Clustering

After computing the persistence of Vietoris-Rips filtration, the topological information is encoded in a multiset in the extended plane $\bar{\mathbb{R}}^2$, i.e., the persistence diagram. To measure the distance between persistence diagrams, we define what is called the bottleneck distance. Suppose $v = (v_1, v_2)$ and $u = (u_1, u_2)$ are two points in $\bar{\mathbb{R}}^2$. Then $||u - v||_\infty$ is defined to be the maximum of $|v_1 - u_1|$ and $|v_2 - u_2|$. Let P_1 and P_2 be two multisets. The bottleneck distance between P_1 and P_2 is

$$d_{BN}(P_1, P_2) = \inf_{\gamma} \sup_{u \in P_1} ||u - \gamma(u)||_\infty. \tag{8}$$

where $\gamma : P_1 \to P_2$ ranges over all the bijections between P_1 and P_2. Bottleneck distance satisfies the positivity, symmetry and triangle inequality rules. Therefore, it defines a metric on the space of persistence diagrams. Now that the bottleneck distance matrix is obtained, we apply the standard kmeans clustering algorithm to classify all the persistence diagrams. Since each persistence diagram is the topological feature of sound signals, the kmeans algorithm classify the sound signals as well.

IV. Experiments

We apply our algorithm to classify animal sound data. We download one elephant sound, two bee sounds, two nightingale sounds and three lion sounds. The data can be found in https://github.com/YueqiCao/ESC. We use time-delay embedding map to transform the sound signals into point clouds. The embedding dimension is set to be 10 and the delay is set to be 1. The eccentricity map is used to obtain a subsample. The persistence diagrams and bottleneck distances are computed by Python using the GUDHI module [25]. Visualizations of point clouds and the persistence diagrams are presented in figure 2. From the first two columns we see that the time-delay embedding map gives different topology of point clouds with respect to different signals. The last two columns show that the differences are measured by the bottleneck distances of persistent diagrams. The clustering result is presented after multidimensional scaling (MDS). From figure 3, we see that our classification algorithm successfully classifies the sound signals.

V. Conclusions

This paper introduces a new method to classify the environment sound signals. Unlike any other traditional learning methods, we focus on the topological features of point clouds after time-delay embedding. Bottleneck distance is used to measure the difference of persistence diagrams and support the kmeans clustering. Experiments are carried out to validate the efficiency of our algorithm.

More experiments will be carried out in the future. It is also natural to apply our method to other signal classification problems. We believe there are interesting results in future work.

References

[1] S.Chachada, and C.C.J.Kuo, "Environmental sound recognition: A survey," APSIPA Transactions on Signal and Information Processing, vol. 3, 2014.

[2] S.Duan, J.Zhang, P.Roe, and M.Towsey, "A survey of tagging techniques for music, speech and environmental sound," Artificial Intelligence Review, vol. 42(4), pp. 637–661, 2014.

[3] N.Yamakawa, T.Takahashi, T.Kitahara, T.Ogata, and H.G.Okuno, "Environmental sound recognition for robot audition using matching-pursuit," International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems , Springer, Berlin, Heidelberg, pp. 1–10, June, 2011.

[4] M.Vacher, F.Portet, A.Fleury, and N.Noury, "Challenges in the processing of audio channels for ambient assisted living," The 12th IEEE International Conference on e-Health Networking, Applications and Services, pp. 330–337, July, 2010.

[5] J.C.Wang, J.F.Wang, K.W.He, and C.S.Hsu, “Environmental sound classification using hybrid SVM/KNN classifier and MPEG-7 audio low-level descriptor,” The 2006 IEEE International Joint Conference on Neural Network Proceedings, pp. 1731–1735, July, 2006.
[6] X.Anda, A.Assadi, and Y.Cao, “Calculating Persistence Homology of Attractors in High-Dimensional Chaotic Systems,” 2017 International Conference on Computational Science and Computational Intelligence (CSCI), pp.158–160, 2017.
[7] G.Muhammad, Y.A.Alotaibi, M.Alsulaiman, and M.N.Huda, “Environment recognition using selected MPEG-7 audio features and Mel-Frequency Cepstral Coefficients,” 2010 Fifth International Conference on Digital Telecommunications, pp. 11–16, June, 2010.
[8] G.Carlsson, “Topology and data,” Bulletin of the American Mathematical Society, vol. 46(2), pp. 255–308, 2009.
[9] G.Carlsson, T.Ishkhanov, V.De Silva, and A.Zomorodian, “On the local behavior of spaces of natural images,” International journal of computer vision, vol. 76(1), pp. 1–12, 2008.
[10] G.Carlsson, A.Zomorodian, A.Collins, and L.J.Guibas, “Persistence barcodes for shapes,” International Journal of Shape Modeling, vol.11(02), pp. 149–187, 2005.
[11] J.A. Perea and J. Harer, “Sliding windows and persistence: An application of topological methods to signal analysis,” Foundations of Computational Mathematics, vol.15(3), pp.799–838, 2015.
[12] F.Chazal, D.Cohen-Steiner, L.J.Guibas, F.Mmoli, and S.Y.Oudot, “Gromov-Hausdorff stable signatures for shapes using persistence,” Computer Graphics Forum, Vol. 28, No. 5, pp. 1393–1403, July, 2009.
[13] D.Cohen-Steiner, H.Edelsbrunner, and J.Harer, “Stability of persistence diagrams,” Discrete & Computational Geometry, vol. 37(1), pp.103–120, 2007.
[14] L.M.Seversky, S.Davis, and M.Berger, “On time-series topological data analysis: New data and opportunities,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 59–67, 2016.
[15] S.Emrani, H.Chintakunta, and H.Krim, “Real time detection of harmonic structure: A case for topological signal analysis,” 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3445–3449, May, 2014.
[16] Y.Umeda, “Time series classification via topological data analysis,” Information and Media Technologies, vol.12, pp.228–239, 2017.
[17] A.Zomorodian, and G.Carlsson, “Computing persistent homology,” Discrete & Computational Geometry, vol. 33(2), pp.249–274, 2005.
[18] H.Edelsbrunner, and J.Harer, “Computational topology: an introduction,” American Mathematical Soc, pp. 67–74, 2010.
[19] N.H.Packard, J.P.Crutchfield, J.D.Farmer, and R.S.Shaw, “Geometry from a time series,” Physical review letters, vol.45(9), pp.712, 1980.
[20] F.Takens, “Detecting strange attractors in turbulence,” Dynamical systems and turbulence, pp. 366–381, 1981.
[21] H.Krim, T.Gentimis, and H.Chintakunta, “Discovering the whole by the coarse: A topological paradigm for data analysis,” IEEE Signal Processing Magazine, vol.33(2), pp.95–104, 2016.
[22] S.Emrani, T.Gentimis, and H.Krim, “Persistent homology of delay embeddings and its application to wheeze detection,” IEEE Signal Processing Letters, vol.21(4), pp.459–463, 2014.
[23] R.Ghrist, and A.Muhammad, “Coverage and hole-detection in sensor networks via homology,” Proceedings of the 4th international symposium on Information processing in sensor networks, pp.34, April, 2005.
[24] D.Cohen-Steiner, H.Edelsbrunner, and J.Harer, “Stability of persistence diagrams,” Discrete & Computational Geometry, vol.37(1), pp.103–120, 2007.
[25] The GUDHI Project, “GUDHI User and Reference Manual”, http://gudhi.gforge.inria.fr/doc/latest/, 2015.